DOCUMENT RESUME

ED 163 088 TM 008 229

AUTHOR Evans, Franklin R.
TITLE The GRE-Q Coaching/Instruction Study. GRE Board Professional Report GREB No. 71-5aP.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE Sep 77
NOTE 34p.; Not available in hard copy due to print quality
AVAILABLE FROM Graduate Record Examinations, Educational Testing Service, Princeton, New Jersey 08541 (free while supplies last)

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Aptitude Tests; Black Students; Caucasian Students; *College Entrance Examinations; Cultural Differences; Graduate Study; Higher Education; Instructional Programs; *Mathematics; Mexican Americans; *Program Effectiveness; Racial Differences; Sex Differences; Testing Problems; *Test Wiseness; Tutorial Programs; *Tutoring
IDENTIFIERS *Graduate Record Examinations

ABSTRACT

A special course was designed to aid undergraduate subjects in preparing for the Graduate Record Examination--Quantitative (GRE-Q). The course included a short one-session discussion (Anxiety Reduction Session) of the GRE and its uses, and four sessions devoted to specific instruction in the basic mathematics required for the test and strategies for approaching the various types of questions. The content outline of the course is included. The course was offered to junior-year volunteers at 12 colleges. Colleges were selected so as to maximize the participation of black and Chicano students; predominantly white schools were also included. Detailed results are presented; however, the reader is cautioned that the differences observed in the analyses, while statistically significant, may be due to factors not under the control of the investigator such as differential attrition rates and levels of motivation for various subject groups. The results of the analyses suggested that: (1) there was a small, consistent increase in GRE-Q scores due to the overall program; (2) the increase appeared to occur early in the program; and (3) there was no evidence that the program was differentially effective for the sexes or the ethnic groups studied. (Author/ROF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

THE GRADUATE RECORD EXAMINATIONS

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Educational Testing Service
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED163088

TM008 229

GRE

THE GRE-Q COACHING/INSTRUCTION STUDY

Franklin R. Evans

BEST COPY AVAILABLE

GRE Board Professional Report GREB No. 71-5aP

September 1977

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY ◻ BERKELEY, CALIFORNIA ◻ EVANSTON, ILLINOIS

THE GRE-Q COACHING/INSTRUCTION STUDY

GRE No. 71-5aP

Franklin R. Evans

Copyright © 1977 by Educational Testing Service. All rights reserved.

THE GRE-Q COACHING/INSTRUCTION STUDY

Until recently, the conclusions of research studies designed to test whether some type of short-term instruction[1] could produce meaningful gains on standardized aptitude test scores have been uniformly negative. In these studies, several of which have been conducted by or for the College Entrance Examination Board (CEEB) since 1950, a number of problems are evident. These problems are based upon the subjects' level of ability and test sophistication, the amount of instruction provided, and the quality of the instructional materials and of the instruction itself. In most studies, the subjects tended to be academically very able and highly sophisticated in test-taking strategies, and they were therefore not likely to benefit from short-term instruction. A more recent study (Roberts & Oppenheim, 1966), concerned with disadvantaged students, indicated that their subjects were probably too deficient in their general academic preparation to benefit from short-term instruction for an SAT-level test. In most of the studies that have been conducted, instruction was too brief (from 7 to 15 hours at most) and too diverse (including both verbal and mathematics content) to test the effectiveness of short-term instruction. Finally, the instruction was typically unstructured, no attempt having been made to develop a systematic set of instructional materials to be administered by well-trained and properly oriented teachers.

The question of the susceptibility of the SAT to short-term intervention arose again when it was proposed that a new type of mathematics aptitude item (Quantitative Comparison, or QC) replace the more traditional (Regular Math, or RM) and the Data Sufficiency (DS) items currently found in the SAT-M. The new item format had the demonstrated advantages of providing greater reliability and apparently equivalent validity for a given unit of testing time, but there was a concern about the susceptibility of the new item type to short-term intervention.

---

[1] Throughout this report, the terms short-term instruction and short-term intervention are used interchangeably to mean a structured program including instruction for content as well as test-taking strategies. Such instruction, if successful, would probably result in changes in criterion performance as well as test performance. Coaching is traditionally defined as instruction in test-taking strategies and/or in the peculiarities of a particular test. Coaching, if successful, would probably result in changes in test performance, but not in criterion performance.

At the request of the CEEB, Evans and Pike (1973) undertook a study to determine the relative susceptibility of the three mathematics item formats (QC, DS, and RM) to score changes resulting from short-term intervention. The main emphasis of the SAT-M study was to determine whether the QC item format was susceptible to some sort of short-term intervention and, if so, to compare its susceptibility with that of the RM and DS formats.

In general, the results of the Evans and Pike study confirmed the hypothesis that each of the three item formats was at least somewhat susceptible to instruction. The study clearly established that if highly motivated high school students are given well-prepared and well-presented short-term instruction in appropriate mathematics content and test strategies, significant gains in mathematics aptitude test scores are likely to result.

Given the success of this study, it seemed likely that similar results could be obtained with subjects taking aptitude tests at a higher level--that is, candidates for the Graduate Record Examinations. The GRE-Q is similar to the SAT-M in general content and difficulty level. Neither test assumes that candidates have received instruction in mathematics beyond high school algebra and geometry, and many of the candidates are neither highly sophisticated nor necessarily "fresh" in their understanding of basic mathematics concepts. There are several differences, however. The SAT-M at that time used two item formats, Regular Math and Data Sufficiency, with a fairly large percentage of geometry items. The GRE-Q does not use Data Sufficiency items, but does, on the other hand, use the Data Interpretation format that requires the candidate to answer questions using data presented in tabular or graphical form. It also contains very few geometry items and has many more word problems than the SAT-M.

The purposes of the present study were to determine the susceptibility of the GRE-Q to short-term instruction and to assess the differential effectiveness of this instruction for black, white, and Chicano GRE candidates. The study included four phases: (1) the feasibility study, (2) the exploratory study, (3) the developmental phase, and (4) the operational phase.

The Feasibility Study

The feasibility study was a survey of several colleges similar to those that would be asked to participate in the final operational phase of the study. The purpose of this survey was to determine the extent of students' interest in and need for a

program of special intensive instruction for the mathematical part of the Graduate Record Examination, and to determine the degree of interest and cooperation of the faculty and administration in the program.

The Sample

Colleges. Twelve colleges, chosen because of their geographic location and the nature of their student bodies, were invited to participate in the feasibility study. Two geographic areas were identified for this sample: the Southeast/Southwest region and the Northeast/Midwest/Far West region. A distribution in the range of colleges according to the nature of their student bodies was sought on the basis of racial or ethnic composition: integrated (multiracial) or predominantly black, Chicano, or white. Two of the 12 schools (both integrated) originally invited declined to participate in the survey. Of the 10 remaining schools, four were predominantly black, two predominantly Chicano, two predominantly white, and two integrated.

Students. Because of time pressures, the selection of students to be surveyed was left to the discretion of the college administrator who conducted the survey on each campus. Some colleges chose to survey a one-half or one-third sample of their junior classes. In these colleges, a questionnaire was mailed to each student chosen for the survey, and the student returned the completed questionnaire directly to ETS. In other colleges where juniors were surveyed in the classroom, the questionnaires were collected in the classroom and returned to ETS in one batch. The principal investigator conducted the survey on two of the campuses. On these campuses administration of the survey was preceded by a lecture about the GRE and its various items. Although the lectures were advertised and were scheduled at a time when they would not conflict with other campus activities, on both campuses conflicts with other important activities did arise and probably accounted for the somewhat smaller than expected attendance. At each of these campuses additional responses to the survey questionnaire were solicited through classroom instructors.

Questionnaire

The survey material included a one-page description of the program of instruction and a one-page questionnaire. The description of the program stressed the reasons for conducting a special intensive instructional course for the mathematical portion of the GRE, the nature and content of the course and of the instructional materials, the amount of time involved in classroom instruction and in out-of-class work, and the fact that the course would be provided at no charge to the student. The questionnaire requested

the student to indicate the extent to which he or she would participate in a program such as the one described, and the amount of time per week he or she would be willing to devote to the course, and also asked for information on the student's undergraduate major, intended field of graduate study, and ethnic background.

Results

The results of the feasibility survey for five colleges (two predominantly Chicano, one predominantly white, one predominantly black, and one integrated) are summarized in Table 1. It is unfortunate that many schools were in the midst of final examinations at the time the survey was conducted, and so there was a much lower response rate than had been anticipated. Some schools declined because of end-of-the-year time pressures, and a few schools did not provide usable results.

The data presented in Table 1 were prepared for an interim report to the GRE Research Committee, for which time data from only five schools were available. The decision to continue the study was based on these data.

The response to the questionnaire was overwhelmingly positive. Students indicated that if a special instruction program for GRE-Q were offered on their campus, they would volunteer, and in addition showed a willingness to devote a substantial amount of time to such a program. The majority of the respondents were willing to devote at least two classroom hours per week to the course. All administrators surveyed indicated keen interest in the project and a willingness to cooperate. Discussions with administrators at colleges in the survey that had not yet provided data seemed to support the results shown in Table 1. On the basis of the evidence that sufficient student and administrator interest did exist in the proposed programs, a recommendation to continue the developmental efforts was made to the GREB Research Committee, and the Committee concurred with the recommendation.

The Exploratory Study

The exploratory study was designed to provide the principal investigator and the test development staff with specific information about the problems that students encounter when responding to GRE-Q items, thereby identifying the weakness of students and the areas where intensive instruction was needed. This information then served as a basis for the development of the experimental curriculum. The principal investigator and the test development staff member who was to be involved in developing the instructional materials

Table 1

Summary of Results of Student Survey at Five Colleges

College	Type	Region	Method	Black		Chicano		White		Other ethnic group*		Ethnic group not indicated	
				would participate	would not participate	would participate	would not participate	would participate	would not participate	would participate	would not participate	would participate	would not participate
A	Predominantly Black	SE/SW	Visit-Classroom Survey	85	4	8	2	5	1	0	0	1	0
B	Predominantly White	SE/SW	Mail	6	0	0	0	191	64	4	0	1	1
C	Predominantly Chicano	SE/SW	Visit--Classroom Survey	5	1	98	27	64	16	0	0	0	0
D	Predominantly Chicano	SE/SW	Mail	1	0	16	1	3	3	4	0	0	0
E	Integrated	NE/MW/FW	Mail	31	3	2	0	198	54	12	2	2	2
			Total 918	128	8	124	30	461	138	20	2	4	3

* Not black, Chicano, or white

visited three campuses (one predominantly black, one predominantly Chicano, and one multiracial) and tested and interviewed a total of 44 student volunteers. The volunteers were assembled in a group on each campus and given one of two forms of a 20-minute multiple-choice test consisting of 15 items taken from the GRE-Q item pool. After the test, the research team held a 30-minute interview with each volunteer. During this interview, the student was asked to reconstruct the reasoning process by which he or she had arrived at an answer for each problem on the test. The interviews were taped, and extensive notes were taken. Subsequently, each of the 30 test items was categorized according to the specific mathematical concepts considered necessary to answer it.

Responses to the items were then categorized according to the difficulty that the various concepts presented to the students. From this information, an outline of mathematics concepts that needed to be covered in an intensive instructional program was developed. That outline is presented in Table 2. The research staff and the test development staff refined the outline and devised plans to develop materials (teachers' outlines and student workbooks) to cover each of the major topics of the outline. The original plan of the study called for six classroom sessions to cover the specified mathematical concepts. The GREB Research Committee requested, however, that an attempt be made to test the effectiveness of a two-hour test familiarization session. Because of the additional time necessary for that session and for post-testing to measure its effectiveness, the mathematics concepts were covered in four classroom periods rather than six.

The Developmental Phase

The coaching/instruction curriculum materials that were developed required eight sessions of approximately two hours each, the first and last of which were devoted to testing. The second session was devoted to test familiarization aimed at alleviating the candidate's anxiety about the admissions process. For this purpose there was a presentation of a practice with various general testing strategies--e.g., pacing and when and how to guess. Further information included the uses of tests in the admissions process, the legitimacy of admissions tests (validity and reliability), the GRE test directions, and the answer sheets.

The third session consisted of a short (40-item) mathematics test designed as a scaled-down version of the GRE-Q with similar items and content. The purpose of this session was to assess the effects of the second session, and the test was reviewed item by item as a learning exercise.

Table 2

Content Outline

Outline for GRE Math Coaching Study

Lesson 1

Computational Skills

A. Whole numbers

1. Addition and multiplication with carrying
2. Subtraction with borrowing
3. Division
   a. With and without remainders
   b. With powers of 10 (10,000, not 10^4)
   c. Zeros in divisor and dividend (ex. 305)

B. Properties of numbers

1. Odds--evens
2. Factors and divisibility
3. Prime numbers
4. Factorization into primes
5. Least common multiple
6. Greatest common divisor

C. Fractions

1. Reducing and equivalent fractions
2. Multiplication
3. Reciprocals
4. Division
5. Addition and subtraction of unlike fractions

D. Decimal fractions

1. Decimal notation-powers of 10 (not 10^x)
2. Equivalent decimal fractions
3. Addition of decimal fractions
4. Subtraction of decimal fractions
5. Multiplication of decimal fractions

D. Decimal fractions (continued)

6. Division by whole numbers
7. Division by decimal fractions
8. Multiplication by powers of 10
9. Division by powers of 10

Lesson 2

A. Ratio-proportion

B. Percent

C. Averages

Lesson 3

Algebra and geometry, measurement

A. Signed numbers and number lines

B. Evaluation of algebraic expressions

C. Solution of open sentences

1. Equalities and inequalities in 1 unknown
2. 2 equations in 2 unknowns
3. Verbal problems
   a. Setting up
   b. End-points

D. Exponents and roots

1. Basic laws of exponents
2. Square roots (emphasis on squaring, not taking roots)
3. Expanded notation

E. Operations with algebraic fractions

F. Area and perimeter of triangle, parallelogram, rectangle, and square, and the diagonal of the square

G. Area and circumference of a circle

H. Volume of rectangular solid, cube, cylinder, and sphere

Algebra and geometry, measurement (continued)

I. Isosceles and equilateral triangles

J. Pythagorean theorem, right triangles, and special triangles (3-4-5, 30(1) — 60(2) — 90($\sqrt{3}$), 45(1) - 45(1) - 90 ($\sqrt{2}$))

K. Relation between size of sides and angles of a triangle

L. Angle measurement

1. Acute, right, obtuse, straight
2. Vertical
3. Sum of angles of triangle and quadrilateral
4. Angle measurement related to parallel lines

M. Coordinate geometry [rectangular coordinate system]

Lesson 4

Data interpretation

A. Bar graphs

B. Line graphs

C. Circle graphs

D. Tables and charts

E. Unusual graphs [triangular, flowcharts, etc.]

Sessions 4 through 7 covered the basic content that had been identified in the exploratory study as necessary for adequate performance on the GRE-Q. These sessions and the content of each are listed below:

Session 4 - basic number facts, operations with fractions, lowest common multiple, greatest common divisor.

Session 5 - averages, ratio and proportion, and percent.

Session 6 - basic linear algebra, geometry.

Session 7 - data interpretation.

The curriculum materials were developed mainly by the mathematics staff of the ETS Test Development Division, using standard procedures for item-writing and for the preparation of the tests and for the workbooks, class books, mini-lessons, and lesson plans. Detailed explanations of most problems included in the materials allowed the student to follow the solutions step-by-step.

The materials included lesson plans for all sessions for the teachers, and workbooks, class books, and mini-lessons for Sessions 4 through 7 for the students. The lesson plans for the three testing sessions (Sessions 1, 3, and 8) gave instructions on how to administer the tests. The lesson plan for the test familiarization session (Session 2) presented ideas for discussion as well as sample GRE questions. The outline of the remaining four lesson plans covered a discussion of the mini-lesson with examples for the class to try, a discussion of the topics to be covered in that class (again with examples), and keys and hints for the classroom exercises, classroom test, and diagnostic test.

The workbooks contained exercises dealing with the concepts taught in the mini-lessons. The format was one of immediate feedback: the right-hand page contained several problems and space in which the student could work the problems; the reverse side of the page gave explanations of the solution of the problems and the correct answers. The last exercise in each lesson was a timed practice test, with answers and explanations at the end. Students were expected to do the workbook exercises on their own time and to bring up any problems or questions regarding them in the next class.

The class books consisted of a few practice exercises, 5 or 10 minutes long, to be completed by the students in class and then discussed. Each class book also contained a class test and a short diagnostic test. The class test consisted of 15 problems to be solved in 20 minutes. It was suggested that the instructor discuss the class test problems and their solutions in class. The diagnostic test, to be taken in class, was turned in to the instructor for marking after class. The results of the diagnostic test were to be used by the instructor to determine particular strengths and weaknesses of the class.

The mini-lessons were one- or two-page summaries of the major points to be covered in the lessons. They were designed for students to use for review and as reference material. The student was encouraged to memorize the concepts presented in the mini-lessons.

The Operational Phase

The next phase of the study was designed to test the effectiveness of the course of instruction described above as a means of increasing the GRE-Q scores of candidates, especially minority candidates. In order to accomplish this objective, a special course was offered to volunteer GRE candidates on 12 campuses.

Procedures

Schools. The schools invited to participate in the project were selected according to the makeup of the student body, school control (state or private), and geographical location. The schools included four predominantly black schools in the southeast and southwest United States, one predominantly Chicano school in the southwest, two predominantly white schools in the southeast, three multiracial white-black schools in the midwest and northeast, and two multiracial white-Chicano schools in the southwest. Some of the schools were selected on the basis of the results of the feasibility study, and some were selected because of their expressed interest in the project.

Coordinators. An attempt was made to select on each campus a project coordinator who was highly visible and was respected by the target groups. For example, on one campus, the project coordinator was a Chicano man working in the office of Special Projects and dealing directly with the academic problems of minority students. On other campuses, coordinators were connected with academic foundations departments (minorities special studies programs), testing offices, and psychology departments. A coordinator's main role was as a liaison between the project staff in Princeton, the instructors, and the administration of the university. Specifically, coordinators were asked to suggest potential instructors for the courses, to make arrangements for space for testing and instruction, and to recruit student volunteers for the project.

Teachers. Teachers were recruited from the faculties of nearby high schools or from college instructors in special remedial programs. The decision not to use college mathematics professors was based on the belief that such persons would not be extremely interested in or effective at teaching mathematics at this level and of a remedial

nature. Of the 17 teachers engaged for the project, eight were currently teaching mathematics at the high school level, and nine were remedial education specialists at the college level. A one-day workshop was held in August 1973, for the purpose of familiarizing the teachers with the materials and procedures of the project. Sixteen of the 17 teachers attended.

In July 1973, instructions for recruiting and registering student volunteers for the course were sent to the project coordinator. The instructions outlined criteria for student participation and suggested ways to go about recruiting and registering students. The principal criteria for student participation in the project were that students should be seniors, should be thinking of attending graduate school and therefore considering taking the GRE aptitude tests, and should not be mathematics or "hard science" (e.g., physics) majors. The coordinators were instructed to inform students about the experimental nature of the project and the amount of time and work, as well as testing, that would be required. The actual procedures used to recruit students were left to the discretion of the coordinators. They were provided, however, with sample press releases and a sample announcement to mail to each student selected for the project. They also had an outline of the course and detailed instructions for registering the students.

The number of student volunteers to be recruited on each campus depended upon the size and type of school. The seven predominantly uniracial schools had one class for each of the experimental and control groups; the five multiracial schools had two classes for each of the experimental and control groups--a total of 17 classes. The maximum enrollment of each class was 20 students, 10 male and 10 female. On the predominantly uniracial campuses, coordinators were asked to recruit 40 subjects, 20 male and 20 female. Half of each of these groups were to be assigned at random to either an experimental or a control group. On the multiracial campuses, coordinators were asked to recruit 80 volunteer subjects, 40 minority and 40 nonminority, half female and half male, from which the experimental and control groups would be assigned randomly.

Results

Table 3 shows the number of students called for by the original plan on each of the 12 campuses. Table 4 is a diagram of the research design for subjects in each school. The two experimental and two control groups represent random halves of these groups respectively. Table 5 shows the number of students actually recruited on each of the campuses--that is, the number of students who volunteered for the project and who appeared for the initial session. The last column of Table 5 (Complete Data) shows the total number of subjects in each school (without regard to sex or ethnic group) from each of whom a usable pretest and posttest were obtained.

Table 3

Number and Type of Students to Be Recruited
from Each of 12 Schools

School Number	School Type and Control*	Number of Classes	Black M	Black F	White M	White F	Chicano M	Chicano F	Total M	Total F
01	Black (Pr.)	1	20	20	-	-	-	-	20	20
02	Black (St.)	1	20	20	-	-	-	-	20	20
03	Chicano-White (St.)	2	-	-	20	20	20	20	40	40
04	Black-White (St.)	2	20	20	-	-	20	20	40	40
05	Chicano (St.)	1	-	-	-	-	20	20	20	20
06	Chicano-White (St.)	2	-	-	20	20	20	20	40	40
07	Black-White (St.)	2	20	20	20	20	-	-	40	40
08	White (St.)	1	-	-	20	20	-	-	20	20
09	Black (Pr.)	1	20	20	-	-	-	-	20	20
10	White (Pr.)	1	-	-	20	20	-	-	20	20
11	Black (St.)	1	20	20	-	-	-	-	20	20
12	Black-White (St.)	2	20	20	20	20	-	-	40	40
	Totals	17	140	140	120	120	80	80	340	340
									680	

*Pr. = Private; St. = State

Table 4

Diagram of Experimental Design for Subjects in Any School

Group [1]	SESSION							
	1	2	3	4	5	6	7	8
Experimental I	Pretest Form 1 + Form 3a	Test Familiariza-tion	Posttest for Test Familiar-ization. Form 3a + 3b	Content Lesson I: Basic Number Facts	Content Lesson II: Averages, Ratio & Proportion, Percent	Content Lesson III: Linear Algebra, Geometry	Content Lesson IV: Data Interpretation	Posttest Form 5
Experimental II	Pretest Form 2 + Form 3b	Test Familiariza-tion	Posttest for Test Familiar-ization Form 3a + Form 3b	Content Lesson I: Basic Number Facts	Content Lesson II: Averages, Ratio & Proportion, Percent	Content Lesson III: Linear Algebra, Geometry	Content Lesson IV: Data Interpretation	Posttest Form 4
Control I	Pretest Form 1 + Form 3a	Posttest Form 5	Test Familiariza-tion	Posttest for Test Familiar-ization Form 3a + Form 3b	Content Lesson I: Basic Number Facts	Content Lesson II: Averages, Ratio & Proportion, Percent	Content Lesson III: Linear Algebra, Geometry	Content Lesson IV: Data Interpretation
Control II	Pretest Form 2 + Form 3b	Posttest Form 4	Test Familiariza-tion	Posttest for Test Familiar-ization Form 3a + Form 3b	Content Lesson I: Basic Number Facts	Content Lesson II: Averages, Ratio & Proportion, Percent	Content Lesson III: Linear Algebra, Geometry	Content Lesson IV: Data Interpreta-tion

Form 1 + Form 3a = 60 items = 85 minutes
Form 2 + Form 3b = 60 items = 85 minutes
Form 3a + Form 3b = 40 items = 55 minutes
Form 4 = Form 1 = 40 items = 55 minutes
Form 5 = Form 2 = 40 items = 55 minutes

1. The two experimental and two control groups are random halves of the experimental and control groups, respectively.

Table 5

Number of Subjects Registered in Each Ethnic and Treatment Group, Total Potential Subjects, and Subjects Who Supplied Proper Pre- and Posttests at 12 Schools

School	Group	Registered									Potential Subjects	Complete Data
		Black			White			Chicano				
		Male	Female	Total	Male	Female	Total	Male	Female	Total		
01	E	9	9	18							18	4
	C	5	11	16							16	13
02	E	12	13	25	1	9	1				26	0
	C	12	12	24	0	0	0				24	0
03	E	1	0	1	6	13	19	6	0	6	26	5
	C	0	0	0	11	14	25	2	2	4	29	28
04	E	4	3	7	4	6	10	0	1	1	18	10
	C	0	4	4	2	10	12	0	0	0	16	11
05	E				2	1	3	4	7	11	14	8
	C				3	1	4	4	2	6	10	7
06	E	0	0	0	9	9	18	10	8	18	36	25
	C	0	1	1	9	9	18	12	7	19	38	29
07	E	5	8	13	8	9	17	2	0	2	32	9
	C	2	9	11	5	7	12	0	0	0	23	18
08	E				7	9	16				16	7
	C				7	12	19				19	19
09	E	5	12	17							17	10
	C	6	13	19							19	17
10	E	1	2	3	0	4	4				7	5
	C	1	1	2	1	3	4				6	6
11	E	3	11	14				0	2	2	16	1
	C	3	11	14				1	1	2	16	8
12	E	2	6	8	2	3	5	(1)	(1)	(2)	15	5
	C	2	5	7	2	6	8	(0)	(8)	(8)	23	18
Total	E	42	64	106	39	54	93	22(1)	18(1)	40(2)	241	89
	C	31	67	98	40	62	102	19(0)	12(8)	31(8)	239	174

* Numbers in () in these columns represent Puerto Rican students.

It should be noted that two schools (2 and 11) provided no usable data. When sex and ethnic identity are included as design variables, the cell sizes drop to zero in numerous instances. Also "complete data" refers only to students from whom the full pretest (Session 1) and the full posttest (Session 2--control group, and Session 8--experimental group) were obtained. Any attempt to assess the effects of the brief test familiarization program would result in even smaller sample sizes. These data are at best minimal.

Because of the spottiness of the data, the approach to the analysis was to use all of the data available to answer the following three questions.[2]

1. Was the overall program of instruction helpful in increasing GRE-Q scores of the subjects? Was it equally helpful for all ethnic groups and both sexes?

2. Was the effect, if any, of the program due to the 4 sessions of mathematics instruction alone? Is this answer the same for all ethnic groups and both sexes?

3. Was the overall effect simply due to the anxiety reduction session alone? Is this answer the same for both sexes and each ethnic group?

Question 1 was investigated by using data on all experimental group subjects who took both the pretest (PR) and the posttest (PT). Question 2 was investigated by using data from all experimental group subjects on whom both PT and anxiety reduction (AR) test scores were available. Question 3 used data from all control group subjects on whom both PR and AR tests were available. In each case the analysis used every subject (in the appropriate group) for whom the two scores in question were available.

It was expected that the effectiveness of the program would be dependent upon the instructor and other aspects of educational settings as well as upon the student. Therefore, data were examined for each individual class. The underlying approach to the analysis, then, was to employ intraindividual differences and to analyze the data separately for each class. The results of these analyses are presented below.

---

[2]The analyses presented below were suggested by and performed by Lincoln Moses and Suzanna Wong. A complete report of their analyses appears in Brown, B. et al., Biostatistics Case Book Volume One, Stanford, California: Stanford University, 1976. The author wishes to thank them for their assistance.

Question 1

Was the overall eight-week program helpful in increasing the GRE-Q scores of the subjects?

The posttest (PT) minus the pretest (PR) differences for all experimental subjects in each of the 15 classes on whom data were available were computed. The results are presented in Table 6. Since the variations in means across the instructors is large compared with those within instructors, within-instructor variance is disregarded as underestimating the real uncertainty. It can be seen from Table 6 that there is a small consistent positive difference (12 of 15 classes show PT-PR differences in the positive direction). Confidence intervals are shown at the bottom of the table. Both sign test and t-test confidence intervals are presented. When all 15 classes are used, the 96.4% sign-test-based confidence interval shows a significant median effect due to instruction, but the 95% t-confidence interval based on all 15 classes includes zero. However, when the class (14) with only one subject is dropped from the calculations, the 95% t-confidence interval excludes zero. These results indicate an overall positive effect, which is due to the instructional program that included both anxiety reduction and mathematics instruction.

Table 7 presents results related to the differential effects of the instructional program that are due to the sex of the subjects. PT and PR scores were available for both male and female subjects in 13 of the 17 classes. In 6 classes the differences favored females; in the other 7 the differences favored male subjects. Thus, there is no basis on which to conclude that the instruction had a differential effect for either sex.

Similarly, Table 8 presents results related to the differential effects of the instructional program resulting from the subjects' ethnic identities. There were only 5 classes in which Chicano-white comparisons could be made, and in these the differences favored Chicanos in 2 classes and whites in 3 classes. Black-white comparisons could be made in 4 classes with blacks showing greater average gains in one class and lower in 3 classes. These data do not lend support to the hypothesis that the overall instructional program was differentially effective because of the ethnic backgrounds of the subjects. Furthermore, the small numbers of classes available for evaluation would most probably obscure whatever effects might exist.

Question 2

Was the instructional effect due to the mathematics instruction only?

In Table 6 it was demonstrated that there was a significant effect that resulted from instruction. However, the instruction consisted of 2 parts, a one-session program designed to familiarize the subjects with

Table 6

Means and Variances of Intrasubject Differences (PT-PR)

in Experimental Classes

N	Class	Mean	Variance
4	1	0.012581	0.0018412
5	[2* 3]	-0.010213	0.011344
7	[4	0.056568	0.021448
3	5]	0.035356	0.0088164
8	6	0.063622	0.010494
12	[7	0.053932	0.017486
13	8]	0.044184	0.025712
3	[9	0.13959	0.00015771
6	10]	0.077434	0.0031882
7	11	0.034219	0.014671
10	12	0.064124	0.0060382
5	13	0.023538	0.011591
1	[14 15]	-0.17033	
2	[16	0.0078933	0.0056242
3	17]	-0.0028386	0.026072

96.4% sign-test-based confidence interval for median overall instruction effect: (0.0078933 , 0.065415)

95% t-confidence intervals for overall instruction effect:

a. using all 15 instructors (-0.0081273 , 0.065415)

b. omitting case 14 (0.020770 , 0.064942)

c. omitting cases 9 and 14 (0.018664 , 0.052166)

d. omitting the 5 cases with 3 or fewer subjects (0.022680 , 0.061318)

*Bracketed classes are from the same school.

Table 7

Means and Variances for PT-PR Scores for Males and Females in Experimental Classes, Together with Estimated Variances of Those Mean Differences

CLASS	N	MALES Mean	MALES Variance	FEMALES Mean	FEMALES Variance	DIFFERENCE*	V (Diff)**
1	4	0.012581	0.0018412				
[2 ***							
[3	1	-0.097411					
	4			0.011587	0.011957	-0.109000	
[4	3	-0.055680	0.013357				
	4			0.140750	0.011042	-0.196430	0.0074377
[5	1	0.058839					
	2			0.023614	0.168060	0.035225	
6	1	0.123610					
	7			0.055052	0.011558	0.068562	
[7	4	0.054343	0.013361				
	8			0.053727	0.021751	0.0006168	0.0060591
[8	7	0.070853	0.010334				
	6			0.013070	0.047151	0.057784	0.0093347
[9	1	0.138200					
	2			0.140280	0.0003125	-0.002083	
[10	4	0.077260	0.0028801				
	2			0.077781	0.0073003	-0.00052090	0.0043702
11	2	0.12560	0.0094795				
	5			-0.0023343	0.013791	0.12794	0.0074980
12	3	0.12071	0.0099414				
	7			0.039874	0.0034565	0.080835	0.0038076
13	1	-0.097219					
	4			0.053727	0.0093781	-0.15095	
[14	1	-0.17033					
[15							
[16	1	0.060922					
	1			-0.045136		0.10606	
[17	2	-0.066065	0.028158				
	1			0.12361		0.18968	

97.7% sign-test-based confidence interval for median sex difference in overall instruction effect: (-0.15095 , 0.080835)

95% t-confidence interval for sex difference in overall instruction effect: (-0.080861 , 0.054454)

* Male mean minus female mean.

** Estimated variance of mean difference.

*** Bracketed classes are from the same school.

Table 8

Average Scores Differences (PT-PR) for Experimental Subjects Classified by Ethnic Group.

CLASS	WHITE		BLACK		CHICANO	
	N	Mean	N	Mean	N	Mean
1			4	.012581		
3	4	.011587			1	-.097411
4	2	.072477	5	.050204		
5	3	.035356				
6	2	-.009815			6	.088102
7	2	.096531			7	.023505
8	8	.034171			5	.060204
9	3	.13959				
10	4	.069448	1	.14653	1	.040281
11	7	.034219				
12			10	.064124		
13	3	.099939	2	-.091065		
14			1	-.17033		
16	1	.060922	1	-.045136		
17	3	-.002839				

the GRE and its uses in order to reduce any anxiety they may have had about it and 4 sessions of mathematics instruction. In the face of an overall effect, we next wish to determine if that effect is because of something beyond that attainable from the anxiety reduction session. All experimental subjects on whom both posttest (PT) and anxiety reduction (AR) test scores were available were used to answer this question. PT minus AR test score differences were computed for each class on which this combination of scores was available for one or more subjects. The results are presented in Table 9.

For the 11 classes represented in Table 9 the average differences between the instructional program posttest and the test following anxiety reduction were positive in 6 and negative in 5. No reliable differences due to the 4-session instructional program alone were detected; the sign-test based 96.4% confidence intervals and the various 95% t-confidence intervals always include zero. The breakdown for sex and ethnic background differences also failed to show differential effects, and those data are not presented here.

Question 3

Was the overall effect due to the anxiety reduction session alone?

Given that there was a significant overall effect that does not appear to be attributable only to the 4 sessions of mathematics instruction, the next question is whether the effect is due to the anxiety reduction session alone. In order to investigate this hypothesis data from control subjects who had both anxiety reduction (AR) and posttest (PT) scores were used.

The difference AR-PT for the control Ss represents the change in score after the anxiety reduction session where the PT is a second test and AR a third test. If this difference were positive, it would suggest that the anxiety reduction session had a positive effect over and above a mere retest (note from Table 4 that in this case PT is a second and the AR a third test). The results of this analysis are presented in Table 10. There were 12 control classes in which one or more students furnished the desired combination of scores. In two classes the differences were negative and in the other 10 the differences were positive. The confidence intervals (both sign-test and t) shown at the bottom of the table exclude zero in all cases, suggesting that the anxiety reduction session did have an effect on scores that was not merely a "second-test" effect. Also it is not likely that this effect could be due to a third testing since no such effect was found for the experimental subjects even after substantial mathematics instruction. Again, breakdowns by sex and ethnic background failed to show differential effects, and those data are not presented.

Table 9

Means and Variances of Intrasubject Differences (PT-AR)

in Experimental Classes

N	Class	Mean	Variance
4	1	-0.10913	0.0071867
	[2 *		
	[3		
7	[4	0.020382	0.023286
4	[5	-0.083176	0.0050465
	6		
12	[7	-0.047036	0.030128
	[8		
3	[9	0.088094	0.00097656
6	[10	0.016219	0.033309
7	11	-0.011415	0.011321
6	12	-0.054445	0.065852
5	13	0.044063	0.012425
	[14		
	[15		
1	[16	0.190515	
3	[17	0.063901	0.019373

93.4% sign-test-based confidence interval for median increase in scores due to coaching: (-0.054445 , 0.063901)

95% t-confidence intervals for increase in scores due to coaching:

a. using all 11 instructors (-0.46918 , 0.068367)

b. omitting case 16 (-0.053779 , 0.039270)

c. omitting the 3 cases with 3 or fewer subjects (-0.073265 , 0.017130)

* Bracketed classes are from the same school.

Table 10

Means and Variances of Intrasubject Differences (AR-PT)

for Control Classes.

N	Class	Mean	Variance
5	1	0.0029531	0.015338
	[2 *		
	[3		
5	[4	-0.0065627	0.0016835
7	[5	-0.010014	0.014715
	6		
15	[7	0.036031	0.013633
	[8		
8	[9	0.10402	0.033342
2	[10	0.093860	0.000019530
16	11	0.057243	0.014732
11	12	0.061373	0.024786
4	13	0.096633	0.010929
2	[14	0.078235	0.017578
	[15		
7	[16	0.089094	0.024264
10	[17	0.040070	0.013054

96.1% sign-test-based confidence interval for median increase in scores due to anxiety reduction: (0.0029531 , 0.093860)

95% $\underline{t}$-confidence intervals for increase in scores due to anxiety reduction:

a. using all 12 instructors (0.027396 , 0.079759)

b. omitting the 2 cases with 3 or fewer subjects (0.016900 , 0.077268)

*Bracketed classes are from the same school.

Discussion

Given these results, two points deserve special attention. First, the analyses described above address only the question of whether certain differences are significant in a statistical sense, and do not allow for an assessment of the magnitude of those differences. The results suggest that a positive effect that is due to the one session devoted to anxiety reduction is found consistently across several classrooms, but that there is no consistent positive effect resulting only from 4 weeks of substantive instruction. The absence of a significant effect from 4 weeks of substantive instruction presents less difficulty for interpretation than does the presence of a positive effect from one session on anxiety reduction or for the combination of anxiety reduction and substantive instruction. How large a difference is meaningful in terms of the purposes of the GRE-Q? An average increase of, say, 50 scaled-score points caused by instruction might suggest that there is cause for concern, whereas while a statistically significant average increase of, say, 10 scaled-score points might cause no alarm.

In order to estimate the magnitude of the score increases due to the overall program of instruction, pretest (PR) and posttest (PT) scores for 88[3] experimental subjects in 10 schools were transformed to T-scores and compared. The standard scores were computed by setting the mean and standard deviation of the pretest corrected-score[4] for the entire subject pool (N = 254 in 10 schools) on whom both PR and PT scores were available to 50 and 10 respectively. These parameters were then used to scale the corrected score means for the 10 schools to the PR scale. The results of the scaling are shown in Table 11. Looking at the "difference" column of Table 11 for the 10 experimental groups, we can see that in most cases the gains from PR to PT were not substantial. The largest gain was + .598 T-score units for school 7. The average gain for the 88 experimental subjects who completed the program was + .266 T-score units. The reliability of the pre- and posttests is unknown, but if we assume them to be between .8 and .9, we would estimate the standard error of a score to be between .32 and .45 T-score units. In most cases, then, the score gains would be less than the standard error of measurement.

The population to which these results might generalize is also cause for caution. Note from Table 5 that only about 37% of the experimental group persisted to the posttest. It is highly probable that these highly motivated (perhaps test anxious) people differ from persons who would volunteer for a coaching program and from GRE candidates in general.

---

[3] The single subject in classroom No. 14 (see Table 6) was omitted from these computations.

[4] Corrected score = (number correct) $\frac{1}{k-1}$ (number incorrect): where k = the number of response options for an item.

Table 11

N's, Means, and Standard Deviations of Standardized Pretest, Posttest, and Difference Scores for Experimental and Control Groups in 10 Schools

School	Group	N	Standardized Scores				
			Pretest		Posttest		Difference[1]
			Mean[2]	S.D.	Mean	S.D.	Mean
01	E	4	42.79	5.54	43.21	6.20	.42
	C	13	44.97	9.09	47.37	7.11	2.40
03	E	5	51.56	9.56	51.09	6.24	.47
	C	28	52.23	10.12	52.37	10.20	.14
04	E	10	51.36	9.29	54.87	8.14	3.51
	C	11	55.25	7.90	55.50	6.57	.25
05	E	8	46.19	6.27	49.85	6.87	3.66
	C	7	50.46	12.76	46.72	14.39	-3.74
06	E	25	48.56	11.35	51.51	12.54	2.95
	C	29	51.85	10.57	51.71	11.71	-.14
07	E	9	50.61	8.41	56.59	8.41	5.98
	C	18	48.34	10.54	49.31	9.54	.97
08	E	7	48.23	4.35	50.52	4.68	2.29
	C	19	54.01	7.99	54.11	10.37	.10
09	E	10	42.04	7.78	45.21	9.47	3.17
	C	17	40.27	6.56	40.46	6.23	.19
10	E	5	61.16	12.44	60.95	12.61	-.21
	C	6	58.83	6.48	57.65	10.05	-1.18
12	E	5	59.32	12.99	58.71	12.64	-.61
	C	18	50.71	11.06	49.88	9.28	-.83
Total	E	88	49.35	10.03	52.01	10.52	2.65
	C	166	50.35	9.86	50.39	10.04	.04

1. Difference = mean pretest standardized score - mean posttest standardized score.

2. The total subject pool (N = 254) pretest mean and standard deviation were set equal to 50 and 10, respectively. These parameters were then used to scale the group means and standard deviations to pretest units.

Pretest corrected score mean = 11.55
Pretest corrected score standard deviation = 7.41

$$\bar{T}_g = \frac{10\ (\bar{X}_g - 11.55)}{7.41} + 50$$

$$sd_{T_g} = \left[\frac{sd_{x_c}}{7.41}\right] 10$$

Follow-up Survey of Students

In November 1973, a follow-up survey of students who initially registered for the course was conducted to determine why the attrition rate had been so high. Letters were sent to all students for whom we had addresses. Enclosed with the letter was a postcard on which the student was to indicate whether he or she was still attending the sessions and, if not, how many sessions had been attended and why attendance had stopped. Additional comments about the course were also solicited. Out of the total of 494 letters that were sent, 40 were returned as undeliverable. Of the 220 students (48.5%) who responded, 137 (62.3%) indicated that they were still attending the sessions, and 83 (37.7% of the respondents) indicated that they were no longer attending the sessions.

Of those who reported that they were no longer attending the sessions, 49.5% had attended two or fewer sessions (25.3% not any sessions, 12.1% only one session, and 12.1% two sessions), 21.7% had attended three, and 9.6% had attended four sessions. Only 14.4% had attended five or six sessions, and 4.8% did not indicate how many sessions they had attended.

The most common reason (53%) cited for dropping out of the course was schedule conflicts. Approximately 14.3% of the respondents felt they did not need the review, and an equal number stopped attending after taking the GRE in October. About 10.8% changed their plans about going to graduate school.

In addition to giving these reasons for dropping out of the course, some students made one or more specific comments about the course, as follows: the course was too difficult (12.1%), there were too many tests (9.6%), and the instructor was not very good (4.8%). On the whole, the percentage of positive comments about the course (31.3%) was higher than the negative comments (13.3%). The additional 55.6% of the students either made no comment, or their comments were not classifiable as positive or negative.

Of the 137 students who were still attending class sessions 82% supplied one or more comments and/or criticisms of the course. In general, their comments about the course were favorable, with 60.6% saying something good about it and only 4.4% saying something negative. About 17.5% of the respondents made no comments, and 17.5% of the comments were unclassifiable. Among the comments and criticisms of the course were the following:

10.2% thought that the course had increased their confidence in taking the GRE and/or had reduced their anxiety about it:

9.0% felt that there was too much testing;

8.0% felt that there should have been more or longer classes because of the quantity of material to be covered;

9.0% felt that scheduling of the classes was poor;

5.1% felt that coaching for the verbal section of the GRE should be included;

2.9% felt that the course was too difficult;

2.2% felt that the course was too simple;

1.5% felt there was not enough class work;

1.5% felt there should have been more individualized instruction;

8.0% made unfavorable comments about instructor;

7.3% made favorable comments about instructor.

Follow-up Survey of Campus Coordinators and Instructors

In addition to the mail survey of students, the campus coordinators and instructors were asked to report their perceptions of the strengths and weaknesses of the GRE-Q Special Instruction Program. All persons were enthusiastic about the program and believed that students benefited from the instruction and the testing, but most were able to point to some specific problems related to the testing materials and instructional sessions.

Testing. The consensus of the instructors and coordinators was that there was too much testing, especially in the early sessions and that many students were discouraged immediately by the number of tests and questionnaires. They also cited as problems the lack of immediate feedback on the tests and an apparent lack of communication about the actual structure of the course. They reported that some students thought that every session would be a testing session and that there would not be any content instruction.

Materials. In general, the coordinators and instructors reported that student response to the course materials was overwhelmingly favorable. The immediate reinforcement of seeing the right answer and a detailed solution of the problems made the workbooks appealing. The explanations were considered to be concise and to the point, and the workbooks presented a good mix of all types of problems. Some reported that an outline of the subject matter of each session would be a useful addition to the mini-lessons.

Instructional Sessions. Most people felt that the course should have been extended in order to cover the material adequately. At some schools the instructors reported that the two-hour class at times stretched into a three- or four-hour class. They thought that two hours was a good length for a class of this nature, but that there should have been more classes, either two per week for the eight-week period or one a week for more than eight weeks.

The varying level of mathematical sophistication among the students in most classes was frequently mentioned by instructors. The students' recent experience with mathematics ranged from "I haven't had any math since high school" to "I just finished a course in advanced calculus (or in statistics)." The more sophisticated students tended to drop out of the course, whereas the poorly prepared students were often lost and not able to follow the workbook explanations. Many instructors suggested that more than one course be developed so that students at different levels could be placed either in a brief review course, a test-taking strategies course, or an intensive basic mathematics course.

Summary and Conclusions

A special course was designed to aid volunteer subjects in preparing for the GRE-Q. The course included a short one-session discussion (Anxiety Reduction Session) of the GRE and its uses, designed to familiarize candidates with the test and thus reduce their anxiety. Also included were four sessions devoted to specific instruction in the basic mathematics required for the test and strategies for approaching the various types of questions. The course was offered to volunteers at 12 colleges. Colleges were selected so as to maximize the participation of black and Chicano students.

Initial response (as indicated by the number of students who volunteered for the program) was good at all schools. However, attrition throughout the program was high, producing spotty data with which to evaluate the effectiveness of the course. Because of the spotty nature of the data, the original data analysis plans were abandoned and an approach to analysis that used all available data was employed. These analyses did not involve direct comparisons of experimental group gains with those obtained by comparable control groups. Three major questions were investigated:

1. Was the overall program of instruction helpful in increasing GRE-Q scores of the subjects?

2. Was the effect of the program due to the 4 sessions of mathematics instruction alone?

3. Was the effect of the program due to the single anxiety reduction session alone?

In addition, an attempt was made to determine if there were any differential effects that were due to the sex and ethnic backgrounds of the subjects.

The reader is cautioned that differences observed in these analyses, while statistically significant, may be due to factors not under the control of the investigator such as differential attrition rates and levels of motivation for various subject groups. The results of the analyses suggested:

1. there was a small, consistent increase in GRE-Q scores due to the overall program;

2. the increase appeared to occur early in the program; and

3. there was no evidence that the program was differentially effective for the sexes or the ethnic groups studied.